Information Sciences and Technology Department

AIT 526:
Natural Language Processing

Course Syllabus
Fall 2021 - Section DL1
Dr. Liao, PhD
dliao2@gmu.edu

For each section, a customized syllabus with information specific to that section will be made available to registered students via the Blackboard Learning System.

I. Course Description

Objectives
This course emphasizes on both theory and practice to build a foundation of Natural Language Processing (NLP). It focuses more on study of rule-based and statistical NLP methods for analyzing textual data. It covers a broad set of fundamental NLP topics that are building blocks of many applications and teaches skills for creating computer programs that analyze, interpret, and even generate human language. Topics include:

- Lexical, syntactic, and semantic elements of language.
- Statistical properties of language.
- Rule-based and data-driven approaches to analyzing and understanding language.
- Applications to real-world problems.

Prerequisites
1) No NLP background needed
2) Familiarity with Python programming, or have Java/C/C++ programming experience.
3) Knowledge of statistics or probability
4) Desirable: machine learning and/or algorithm design

Special Topics Description
This course emphasizes on both NLP theory and industry-ready hands-on technical skills. It includes the following topics but extends their wings of knowledge to some advanced techniques:

- Chatbots, Regular Expression, Finite State Automata (FSA) and Formal Languages
- N-Grams Language Models & Smoothing
- Word Prediction, Sentence Generation, Text Summarization
- Hidden Markov Models, Part-Of-Speech (POS) Tagging
- Lexical Semantics, Word Sense Disambiguation (WSD), Decision Lists
- Vector Semantics, TF-IDF, Document Similarity
- Information Extraction, Named Entity Recognition (NER) & De-Identification
- Information Retrieval, Question Answering
- Text Data Preprocessing and Cleansing, Web Scraping
- Text Data Visualization (World Clouds, Dependencies, NER, etc.)
- Python, NLTK, SpaCy, and more NLP Programming Tools
- NLP Algorithm Design & NLP Project Management
- Some latest state-of-art technologies, such as deep learning, multimodal-AI, etc.
II. Textbook and Required Materials

For this course, we will be using two required textbooks.

- **Textbook JM**
  - *Speech and Language Processing*, by Jurafsky and Martin (2nd edition). Other editions are not a good substitute. ISBN: 9780131873216

- **Textbook NLP with Python**
  - *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*, by Bird, Klein and Loper (1st edition). This book is freely available online

III. Faculty and Staff

AIT526-NLP, Instructor: Duoduo (Lindi) Liao, Ph.D.
Email: dliao2@gmu.edu

AIT526-NLP, Course Coordinator: Ozlem Uzuner, Ph.D.
Email: ouzuner@gmu.edu

AIT526-NLP, GTA: TBD
Email: TBD Virtual Office Hour

IV. Course Learning Activities and Grade Distribution

This course includes programming assignments, lecture reports, a midterm, and a final project. The individual labs are optional for extra credit. All the coursework except the midterm and optional labs need to be done in teams. There is no penalty for late submissions of programming assignments and optional labs before the dates as shown in the class schedule, but on-time submission can earn extra credit.

Your grade will be based on the following breakdown*:

<table>
<thead>
<tr>
<th>Component</th>
<th>Quantity</th>
<th>Percent of Grade</th>
<th>Extra Credit</th>
<th>On-Time Submission (Extra Credit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online Activities (self-intro, team evaluation, etc.)</td>
<td>∞</td>
<td>4%</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Individual Labs</td>
<td>3</td>
<td>15%</td>
<td>No</td>
<td>Total 1.5 points: 0.5 point each</td>
</tr>
<tr>
<td><strong>Programming Assignments</strong></td>
<td>3</td>
<td>30%</td>
<td>3 points</td>
<td>Total 3 points: 1 point each</td>
</tr>
<tr>
<td>** 1 Video Lecture Report 1 Paper Presentation</td>
<td>2</td>
<td>6%</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Midterm (Open-Book/Notes/Internet)</td>
<td>1</td>
<td>20%</td>
<td>2.5 points</td>
<td>No</td>
</tr>
<tr>
<td>**Final Project</td>
<td>1</td>
<td>25%</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td>Up to 110%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Subject to revision before and throughout the course.
** Teamwork (3 or 4)

VI. Course Duration, Class Participation, and Schedule

Course Duration

Dates: GMU Academic Calendar: [https://registrar.gmu.edu/calendars/](https://registrar.gmu.edu/calendars/)
Total Duration: 16 weeks

Class Participation
This course is designed for mixed "synchronous" and "asynchronous" delivery (https://masononline.gmu.edu/course-delivery-methods):

- **Synchronous**: Classes are held on a set schedule, and students virtually attend an instructor-led session on a regular basis.
- **Asynchronous**: Students can study at their own pace, accessing instructional materials online in Blackboard at any time, though a set schedule with due dates is still there.

For this course, there are several **important online meetings** ("Synchronous" mode) as scheduled on Blackboard Collaborate Ultra. All students are required to virtually attend the instructor-led sessions on a regular basis.

**Schedule**

A detailed schedule will be published on Blackboard. As many factors may affect the development and progress of a class, the instructor reserves the right to alter the schedule as may be required to assure attainment of course objectives. The schedule is subject to revision before and throughout the course.

*Registered students should see the Blackboard Learning System for the latest class schedule.*

**VII. Grading Guidelines and Grade Scale**

**Grading Guidelines**

Some grade components are evaluated subjectively
A: consistently above and beyond the course/assignment requirements  
B: meets and occasionally exceeds the course/assignment requirements  
C: minimally meets the course/assignment requirements  
F: fails to meet the course/assignment requirements

Grades will be awarded in accordance with the Mason Grading System for graduate students. See the university catalog for policies: http://catalog.gmu.edu for more information.

**Grading Scale**

- The grading scale for this course, is:

<table>
<thead>
<tr>
<th>Raw Score</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>97 – 100%</td>
<td>A+</td>
</tr>
<tr>
<td>93 – 96%</td>
<td>A</td>
</tr>
<tr>
<td>90 – 92%</td>
<td>A-</td>
</tr>
<tr>
<td>87 – 89%</td>
<td>B+</td>
</tr>
<tr>
<td>83 – 86%</td>
<td>B</td>
</tr>
<tr>
<td>77 – 82%</td>
<td>B-</td>
</tr>
<tr>
<td>70 – 76%</td>
<td>C</td>
</tr>
<tr>
<td>0 – 69%</td>
<td>F</td>
</tr>
</tbody>
</table>

**NOTE:** **Study success takes constant effort!**

Instructor will **double check** all students’ coursework graded by GTA throughout the course and also select some assignments to grade directly.

Raw scores may be **adjusted** by the Instructor to calculate final grades.
Students are responsible for checking the currency of their grade books. Grade discrepancies must be brought to instructor’s attention within one week of assignment submission and 48 hours of exam submission.

No make-up for any activity, unless arranged in advance. Only in special cases, such as medical problems and family emergency, make-ups and late assignments may be allowed with verifiable proof.

Final grades will be posted to PatriotWeb, which is the only vehicle for students to obtain those grades. A student with a "hold" on his/her PatriotWeb account will be unable to access final grades until the hold has been removed by the Registrar.

VIII. Communication, Writing and Submissions

Communication: Course announcements will be made through BlackBoard.

Writing: All discussions, and assignments for this course must be in standard English. Do not use slang or texting abbreviations (i.e., lol). Capitalize and use complete sentences in your discussion responses and in your paper. You can use bulleted lists if they make sense as a way to convey the information. Emoticons are acceptable as long as they are not overused and help with communication.

Before submitting work, be sure to proof read your writing and make sure that any references that you include are correct.

Submission of Work: All work for this class must be submitted as the assignment states.

ASA Style Guide: ASA Style Guides are easy to locate using an internet search. The following link is one that should work well for this class, you can access it by clicking here:

VI. Academic Honesty
An important component in learning is taking on tasks, assignments and exams in an honest effort to do your best possible work. You are expected to turn in and do original work.

VII. Etiquette and Disabilities
Please observe proper “etiquette” and “netiquette” – courteous and appropriate forms of communication and interaction – within this course. This means no personal attacks, obscene language, or intolerant expression. All viewpoints should be respected.

Giving Feedback: This course is designed along the principles of synergy and collaborative learning. Therefore, it is important that all students understand how to provide quality feedback to their peers. Here are a few tips for providing, positive, constructive, and useful feedback to peers.

- Be empathetic and remember that this environment is a safe place for making mistakes
- Use nonjudgmental language and phrases that do not attack an individual. One way of doing this is to ask the individual to discuss his/her process for making the final decision.
- Use specific questions, examples, and references as a way of making your point.
• Make your feedback useful by providing suggestions that the individual can understand and use to improve her/his work.

Disabilities: Please message me if you have a disability so we can discuss ways to help you succeed in the course. If you need accommodations that would affect the terms of this syllabus, you will need to provide documentation of your disability.